Spatial Data Methods and Vague Regions: A Rough Set Approach

Theresa Beaubouef
Computer Science Department
Southeastern Louisiana University
Hammond, LA 70402 USA
Email: tbeaubouef@selu.edu

Frederick E. Petry and Roy Ladner
Naval Research Laboratory
Mapping, Charting and Geodesy
Stennis Space Center, MS 39529 USA
Email: (fpetry, rladner)@nrlssc.navy.mil

Abstract: Uncertainty management has been considered essential for real world applications, and spatial data and geographic information systems in particular require some means for managing uncertainty and vagueness. Rough sets have been shown to be an effective tool for data mining and uncertainty management in databases. The 9-intersection, region connection calculus (RCC), and egg-yolk methods have proven useful for modeling topological relations in spatial data. In this paper, we apply rough set definitions for topological relationships based on the 9-intersection, RCC, and egg-yolk models for objects with broad boundaries. We show that rough sets can be used to express and improve on topological relationships and concepts defined with these models.

Keywords: spatial data, rough sets, data mining, spatial relationships, vague regions, uncertainty

Introduction

Spatial databases and qualitative reasoning about spatial data are active topics of research encompassing areas such as artificial intelligence, databases and information systems, data mining, and computational geometry. Results from work on spatial data and reasoning are especially useful in geographic information systems (GIS), spatial databases containing data that is geo-referenced to specific locations on the earth, along with mechanisms for reasoning about this spatial data [1,2]. As with any system that attempts to model some aspects of the real world, there must be some mechanism for the management of uncertainty. It has been continually recognized that uncertainty management is particularly necessary in resolving a myriad of problems inherent in spatial information systems [3,4].

Report Documentation Page

Form Approved
OMB No. 0704-0188

Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
09 JUL 2003	**Journal Article (refereed)**	**09-07-2003 to 18-01-2006**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Spatial Data Methods And Vague Regions: A Rough Set Approach	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER **0602435N**

6. AUTHOR(S)	5d. PROJECT NUMBER
Theresa Beaubouef; Frederick Petry; Roy Ladner	5e. TASK NUMBER
	5f. WORK UNIT NUMBER **74673102**

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)	8. PERFORMING ORGANIZATION REPORT NUMBER
Naval Research Laboratory,Marine Geosciences Division,Stennis Space Center,MS,39529-5004	**NRL/JA/7440--03-1010**

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
Office of Naval Research, 800 N. Quincy Street, Arlington, VA, 22217	**ONR**
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT
Uncertainty management has been considered essential for real world applications, and spatial data and geographic information systems in particular require some means for managing uncertainty and vagueness. Rough sets have been shown to be an effective tool for data mining and uncertainty management in databases. The 9-intersection, region connection calculus (RCC), and egg-yolk methods have proven useful for modeling topological relations in spatial data. In this paper, we apply rough set definitions for topological relationships based on the 9-intersection, RCC, and egg-yolk models for objects with broad boundaries. We show that rough sets can be used to express and improve on topological relationships and concepts defined with these models.

15. SUBJECT TERMS
spatial data, rough sets, data mining, spatial relationships, vague regions, uncertainty

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**		**28**	

A spatial database is a collection of data concerning objects located in some reference space, which attempts to model some enterprise in the real world. The real world abounds in uncertainty, and any attempt to model aspects of the world should include some mechanism for incorporating uncertainty. There may be uncertainty in the understanding of the enterprise or in the quality or meaning of the data. There may be uncertainty in the model, which leads to uncertainty in entities or the attributes describing them. And at a higher level, there may be uncertainty about the level of uncertainty prevalent in the various aspects of the database. An ontology for spatial data has been developed in which the terms imperfection, error, imprecision and vagueness are organized into a hierarchy [5]. At the lowest level of vagueness modeling approaches for spatial data are considered including fuzzy set and rough set theory. Furthermore, in spatial data mining applications, one must not only be aware of this uncertainty, but also to exploit it in an effort to discover relationships in the data that might not have been discovered otherwise.

A fundamental aspect of spatial data requiring uncertainty management is topology. Included in topology are relationships between various spatial data entities. Of particular interest are topological relations associated with regions having indeterminate, vague, or otherwise uncertain boundaries.

In relational databases it has been demonstrated that uncertainty may be managed via rough set techniques by incorporating rough sets into the underlying data model [6] and through rough querying of crisp data [7]. In a previous work [8], we pointed out those areas peculiar to spatial databases and GIS that are in need of uncertainty management and suggested ways in which rough sets techniques may be used to alleviate the problems to result in a better overall system.

In this paper we focus on the problem of uncertainty in topological structures in spatial data, and in particular, to spatial regions with uncertain, broad, or otherwise indeterminate boundaries. An indeterminate boundary may involve uncertainty related to the position of the object, or it may be an intrinsic property of the object itself. Consider, for example, “the Midwest” in the United States. Most people would say that they know where the Midwest is. However, where does the Midwest stop being the Midwest? It does not have a clear-cut boundary, but rather a somewhat vague boundary.

Several methods have been proposed for managing uncertainty associated with vague spatial regions. Fuzzy set approaches have been frequently used [9-11], and more recently rough sets [12,13]. In this paper we investigate the application of rough sets [14] for expressing binary topological relations of the 9-intersection model proposed by [15] and extended for regions with broad boundaries in [16]. We also investigate the application of rough sets for improving the RCC [17,18] and egg-yolk [19,20] models for regions with indeterminate boundaries. We show that spatial relationships expressed with any of these methods can be uniformly expressed by rough sets. We further show that rough sets can represent some types of spatial uncertainty that cannot be expressed using these other methods.

Background: Rough Sets and Uncertainty in Data

Rough set theory, introduced by Pawlak [14] and discussed in greater detail in [21,22], is a technique for dealing with uncertainty and for identifying cause-effect relationships in databases as a form of data mining and database learning [23]. It has also been used for improved information retrieval [24] and for uncertainty management in relational databases [6,7].

Rough sets involve the following:

U is the *universe*, which cannot be empty,
R is the *indiscernibility relation*, or equivalence relation,
A = (U,R), an ordered pair, is called an *approximation space*,
$[x]_R$ denotes the equivalence class of *R* containing *x*, for any element *x* of *U*,
elementary sets in A - the equivalence classes of R,
definable set in A - any finite union of elementary sets in A.

Hence, the approximation space A results when an equivalence relation R is imposed upon the universe U. This partitions U into equivalence classes called elementary sets that may be used to define other sets in A. A rough set X is then defined in terms of the definable sets in A by the following:

lower approximation of X in A is the set $\underline{R}X = \{x \in U \mid [x]_R \subseteq X\}$

upper approximation of X in A is the set $\overline{\mathrm{R}}\, X = \{x \in U \mid [x]_R \cap X \neq \varnothing\}$.

We may also describe the set approximations in terms of regions. Given the upper and lower approximations $\overline{R}X$ and $\underline{R}X$, of X, the R-positive region of X is $POS_R(X) = \underline{R}X$, the R-negative region of X is $NEG_R(X) = U - \overline{R}X$, and the boundary or R-borderline region of X is $BN_R(X) = \overline{R}X - \underline{R}X$. X is called *R-definable* if and only if $\underline{R}X = \overline{R}X$. Otherwise, the lower and upper approximation regions are not equal, and X is *rough* with respect to R. In Figure 1 the universe U is partitioned into equivalence classes denoted by the rectangles. Those classes in the lower approximation of X, $POS_R(X)$, are denoted with the letter P and classes in $NEG_R(X)$ by the letter N. All other classes belong to the boundary region of the upper approximation.

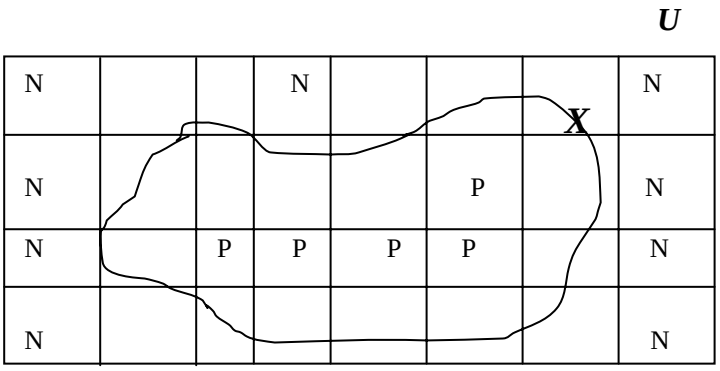


Figure 1. Example of a rough set X.

Consider the following example:

Let U = {BRIDGE, ROAD, STREET, AVENUE, FACTORY, PLANT,
MALL, SHOPS, AIRPORT, FIELD, MEADOW}.

Let the equivalence relation R be defined as follows:

R* = {[BRIDGE], [ROAD, STREET, AVENUE], [FACTORY, PLANT],
[MALL, SHOPS], [AIRPORT], [FIELD, MEADOW]}.

Given some set X = {BRIDGE, ROAD, STREET, AVENUE, FACTORY, MALL}, we can define it in terms of its lower and upper approximations:

$\underline{R}X$ = {BRIDGE, ROAD, STREET, AVENUE}, and

$\overline{R}X$ = {BRIDGE, ROAD, STREET, AVENUE, FACTORY, PLANT, MALL, SHOPS}.

Rough sets, therefore, provide an indiscernibility relation to partition domains into equivalence classes and approximation regions to allow the distinction between certain and possible (or partial) inclusion in a rough set.

The indiscernibility relation allows for the grouping of items based on some definition of 'equivalence' as it relates to the application domain. We may use this partitioning to increase or decrease the granularity of a domain, to group items together that are considered indiscernible for a given purpose, or to "bin" ordered domains into range groups. In data mining applications, this partitioning of domains is varied in systematic ways in an attempt to discover relationships or rules in the data.

In order to allow *possible* results, beyond the obvious certain results encountered in querying an ordinary spatial database system, we may employ the use of the boundary region information in addition to that of the lower approximation region. The results in the lower approximation region are certain corresponding to exact matches. The boundary region of the upper approximation contains those results that are possible, but not certain. The approximation regions play an important role in the representation of vague regions and topological relationships discussed in a later section.

Many of the problems associated with data are prevalent in all types of databases systems. Spatial databases and GIS contain descriptive as well as positional data. The various forms of uncertainty may occur in both types of data, so many of the issues regarding uncertainty apply to ordinary databases as well. See [6,7] for in-depth discussion of incorporation of rough set uncertainty in (non-spatial) databases. These same techniques, including integration of data from multiple sources [25], time-variant data, uncertain data, imprecision in measurement, inconsistent wording of descriptive data, and the "binning" or grouping of data into fixed categories, may also be employed for spatial contexts [8].

Often spatial data is associated with a particular grid. The positions are set up in a regular matrix-like structure and data values are associated with point locations on the grid. There is a tradeoff between the resolution or scale of the grid and the amount of system resources necessary to store and process the data. Higher resolutions provide greater information, but at a cost of memory space and execution time. Data mining applications using high resolution data may sample it at a lower resolution in an effort to improve

performance or to remove “noise” which can sometimes prevent general relationships from being discovered.

There is always indiscernibility inherent in the process of gridding or rasterizing data. A data item at a particular grid point in essence may represent data near the point as well. This is due to the fact that often point data must be mapped to the grid using techniques such as nearest-neighbor, averaging, or statistics. A spatial data application may have the rough set indiscernibility relation defined so that the entire spatial area is partitioned into equivalence classes where each point on the grid belongs to an equivalence class, for example. If this grid resolution is decreased, the granularity of the partitioning is decreased, resulting in fewer but larger classes.

The approximation regions of rough sets apply when information concerning spatial data regions is calculated or displayed. Consider a region such as an airport. One can reasonably conclude that any grid point identified as AIRPORT that is surrounded on all sides by grid points also identified as AIRPORT is, in fact, a point represented by the feature AIRPORT. However, consider points identified as AIRPORT that are adjacent to points identified as MEADOW. Is it not possible that these points represent meadow area as well as airport area but were identified as AIRPORT in the classification process? Likewise, points identified as MEADOW but adjacent to AIRPORT points may represent areas that contain part of the airport. This uncertainty maps naturally to the use of the approximation regions of the rough set theory, where the lower approximation region represents certain data and the boundary region of the upper approximation represents uncertain data. Spatial database querying and spatial database mining operations based on rough sets can incorporate this type of uncertainty for improved results.

By forcing a finer granulation of the partitioning (increase the grid resolution) a smaller boundary region results. As the partitioning becomes finer and finer, the boundary region becomes smaller. When there is no boundary region, the upper and lower approximation regions are the same, and there is no uncertainty in the spatial data.

The 9-intersection, RCC, and egg-yolk methods, which we discuss in later sections, are approaches for handling regions with uncertain boundaries. They use two levels for outlining the vague boundary,

basically corresponding to the approximation regions of rough set theory. These methods, however, provide no facilities for partitioning the domain into equivalence classes, as done in rough sets via the indiscernibility relation. In fact, Roy and Stell [26] discuss the shortcomings of the egg-yolk method if it were to be applied to a discrete rather than a continuous space. They suggest that the egg-yolk method can be used in a multi-resolution context for a finite level of precision and that an extension to the framework may be appropriate. By varying the partitioning in rough sets we can increase or decrease the level of uncertainty present. This results in changes to the approximation regions that define the rough set representation of a region with indeterminate or vague boundaries. The idea that rough sets can, in fact, improve on other spatial data frameworks by quantifying the uncertainty in terms of varying levels of indiscernibility is part of the motivation behind this approach. There are additional benefits gained through the expressive power of rough set theory and representation over other methods.

Topological Uncertainty in Spatial Data

In GIS or spatial databases, it is often the case that we need information concerning the relative positions or distances of objects. Is object A *adjacent to* object B? Or, is object A *near* object B? The first question appears to be fairly straightforward. The system must simply check all the edges of both objects to see if any parts of them are coincident. This yields the *certain* results. However, often in GIS, data is input either automatically via scanners or digitized by humans, and in both cases it is easy for error in the position of data objects to occur. Therefore, we may also want to have the system check to see if object B is very near object A, to derive the *possible* result. If so, the user could be informed that "it is not certain, but it is possible, that A is adjacent to B."

Assume we are investigating coastal bird feeding habitats and trying to uncover any relationships that might exist between the number of bird sightings and coastal structures. One species of bird may require low flat coastal land for feeding on small shellfish. Other types may feed on insects found near their nesting sites in the sides of cliffs. Suppose that for a particular location the system returns results based on the possibility that a high cliff is adjacent to the sea where birds have been sighted that feed in areas of

flat coastal land. We may then be led to investigate the influence of the tides in the area to determine whether low beaches alongside the cliffs are exposed at low tide.

The concepts of connection and overlap can be managed by rough sets in a comparable manner. Connection is similar to adjacency, but related to vector or line type objects rather than area objects. Two objects are connected if they have a common meeting point on one end of each of the objects. It is very easy for spatial data of this type, especially if the data is from different sources, to not align precisely as may occur in the process of conflation of spatial data [27]. We may then want to also define what would constitute *possible* connection, based on perhaps the distance between the objects and the length and orientation of the linear features. For example, if one road feature varying in curvature, but generally oriented from west to east, ends at some point A, and we find a second road, also oriented in an east/west fashion near its beginning at point B a short distance away from A, we may conclude that possibly these two road features are connected, even though they share no common point.

Overlap can be defined in a manner similar to that of nearness with the user deciding how much overlap is required for the lower approximation. Coincidence of a single point may constitute *possible* overlap, as can very close proximity of two objects, if there is a high degree of positional error involved in the data.

Inclusion is related to overlap in the following way. If an object A is completely surrounded by some object B, perhaps we can conclude certainly that A is included in B, lacking additional information about the objects. Equivalently we can say that B "covers" A. If the objects overlap, then it is *possible* that one object includes the other. Approximation regions can be defined to reflect these concepts as well.

Rough sets, 9-intersection modeling, RCC theory, and egg-yolk approaches are useful for managing the types of uncertainty and vagueness related to topology, a few of which were just briefly discussed. These include concepts such as nearness, contiguity, connection, orientation, inclusion, and overlap of spatial entities.

The 9-Intersection Model

In [16], the original 9-intersection model [15] is extended for regions with broad boundaries to result in the 9-intersection matrix depicted in Figure 2 below. Each of the nine entries in this matrix represents a possible intersection relationship between regions A and B, each having broad boundaries. In this extended model, because the boundaries are broad and thereby two-dimensional, certain geometric conditions that held in the original model are no longer valid due to the nature of the boundaries. These conditions are replaced by a set of less restrictive geometric conditions discussed in greater detail in [16].

$$\begin{bmatrix} A^\circ \cap B^\circ & A^\circ \cap \partial B & A^\circ \cap B^- \\ \partial A \cap B^\circ & \partial A \cap \partial B & \partial A \cap B^- \\ A^- \cap B^\circ & A^- \cap \partial B & A^- \cap B^- \end{bmatrix}$$

Figure 2. A° denotes the interior of a region A, ∂A denotes the boundary and the interior of some region A, and A^- denotes the exterior of a region.

Of the 2^9 possible matrices that could be generated by the 3 x 3 matrix, only 44 of the 2^9 are possible, considering the geometric conditions. Many of these matrices correspond to the eight relationships defined [15] for regions with sharp boundaries depicted in Figure 5: disjoint, meet, overlap, coveredby, covers, inside, contains, and equal. Figure 3 depicts a sample of these matrices, along with their graphic representations.

a. 0 0 1
0 0 1
1 1 1

b. 0 0 1
0 1 1
1 1 1

c. 0 0 1
1 1 1
0 0 1

d. 0 1 0
1 1 1
0 1 1

Figure 3. A sample of 9-intersection matrices for relationships between regions A (dashed line) and B (dotted line).

The relationship between two vague regions A and B is represented by placing a '1' at each of the locations in the 3 x 3 matrix where the condition is true, and by placing a '0' at each matrix position where the condition represented for that position is false. We can also say that a one is placed where each of the nine operations given in Figure 2 produces a nonempty region and a zero if the result of the intersection operation is empty.

Examine the first position in the matrix of Figure 2, for example. This position is denoted by $A^\circ \cap B^\circ$. If the inner boundaries of A and B, denoted A° and B°, have no points in common, they do not overlap at all, and a zero is placed in the top left position of the matrix. This is the case for each of the four sample relationships shown in Figure 3.

Now consider the second position of the 9-intersection matrix, which denotes the relationship $A^\circ \cap \partial B$. This relationship produces a nonempty result whenever *the interior of A and the boundary of B* share some point or points in common. In Figure 3 we can see that for the first three samples the relationship $A^\circ \cap \partial B$ results in empty regions. There are no points in common, so the 9-intersection matrix for each of these samples contains a zero in row one, column two. The fourth sample, however, has overlap between A° and ∂B, so a one is placed in that same position for its 9-intersection matrix.

Clementini and di Felice [16] develop a conceptual neighborhood graph based on clustering the relationships together that are geometrically similar. Each node in the hierarchy depicts a 9-intersection configuration for a relationship, and is connected by an arc to those that differ by only one value in the 9-intersection matrix. These 9-intersection techniques, along with clustering are very useful for managing uncertainty involving regions with indeterminate boundaries.

Rough Set Representation of 9-Intersection Model

Rough sets can also be used for expressing spatial relationships via an extended 9-intersection model as shown in Figure 4 below. Here, the lower and upper approximation regions for rough sets A and B defined on universe U are used to define relationships that are equivalent to those in Figure 3.

$$\begin{bmatrix} \underline{R}A \cap \underline{R}B & \underline{R}A \cap (\overline{R}\,B - \underline{R}B) & \underline{R}A \cap (U - \overline{R}\,B) \\ (\overline{R}\,A - \underline{R}A) \cap \underline{R}B & (\overline{R}\,A - \underline{R}A) \cap (\overline{R}\,B - \underline{R}B) & (\overline{R}\,A - \underline{R}A) \cap (U - \overline{R}\,B) \\ (U - \overline{R}\,A) \cap \underline{R}B & (U - \overline{R}\,A) \cap (\overline{R}\,B - \underline{R}B) & (U - \overline{R}\,A) \cap (U - \overline{R}\,B) \end{bmatrix}$$

Figure 4. The 9-intersection matrix expressed in rough set terminology.

Considering Figure 4, we can determine in the rough set representation the 9-intersection matrix values by placing a one at every matrix location where the intersection results in a non-empty region and a zero otherwise. We see that the lower approximation regions of A and B do not intersect for any of the four samples shown. So for each of these, a zero is placed in the 9-intersection matrix for the condition $\underline{R}A \cap \underline{R}B$. The row 1, column 2 matrix location will again contain zeroes for each of the first three samples since the lower approximation of A and the boundary region of B do not intersect. For the fourth sample, however, $\underline{R}A \cap (\overline{R}\,B - \underline{R}B)$ is not empty due to overlap between the lower approximation of A with the boundary region of B. Therefore, a one is placed in this position of the 9-intersection matrix for the fourth sample.

The relationships defined by rough sets in Figure 4 are equivalent to those found in Figure 3. Rough set approaches have mathematical simplicity and elegance, ease of implementation for spatial databases, and a complete and well-elaborated theoretical formulation. Additionally rough sets can provide enhancements to the 9-intersection model in that they also model the uncertainty that arises from indiscernibility, gridding, or partitioning. There is a formal structure in which the partitioning can be varied in order to increase or decrease the level of uncertainty present, which results in changes to the approximation regions.

RCC-8 Theory of Spatial Regions

RCC-8 theory [17,18] is a qualitative reasoning technique for spatial data based on regions rather than points. For any two simple regions, relationships are defined that may hold between them. The eight base relationships that may hold between two given simple regions in the regional connection calculus (RCC-8) are depicted in Figure 5 below, one and only one of which is valid at any given time for a pair of vague regions. These include: PO (Partially Overlapping), TPP (Tangential Proper Part), NTPP (Non-Tangential Proper Part), EQ (Equal), NTPPI (Non-Tangential Proper Part Inverse), TPPI (Tangential Proper Part Inverse), EC (Externally Connected), and DC (DisConnected).

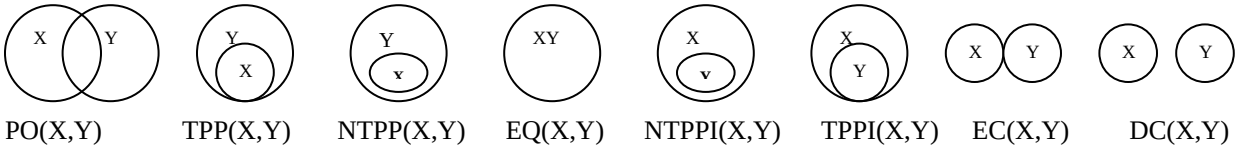


Figure 5. RCC-8 relations.

In [20] the RCC method is extended to apply to regions with vague boundaries rather than simple regions. In that work, only five of the above RCC-8 relations are applicable (called RCC-5). These include PO (Partially Overlapping), PP (Proper Part, when TPP and NTPP are combined), EQ (Equal), PPI (Proper Part Inverse, when NTPPI and TPPI are combined), and DR (Distinct Regions, when EC and DC are combined).

There are a large number of relationships that can occur between two vague regions, since each vague region has two boundaries. The development of [20] lists all the possible relationships and clusters them into a hierarchy based on RCC relations and the effects of "crisping". A vague region X is a refinement, or crisping, of another vague region Y if it can be formed by reducing the imprecision in Y. There are various and incompatible ways of crisping such a vague region. If the imprecision is reduced further, a point will eventually be reached where a crisp, rather than vague, region results. This is called a complete crisping. A sample of the some of the 46 relationships that may exist between two vague regions is depicted in Figure 6.

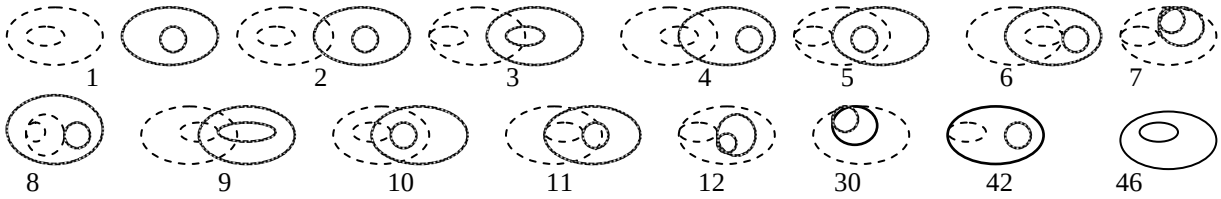


Figure 6: A sample of the 46 possible relationships between regions X (dashed line) and Y (dotted line). A solid line indicates coincidence of an X and Y region boundary. See [11] for the complete listing.

Rough Set Modeling of RCC Relations

Recall that a rough set is comprised of two crisp regions, with each of these regions defined in terms of the underlying equivalence relation. There are many relationships that can hold between two rough regions, each having crisp lower and upper approximation regions.

Distinction between a Single Rough Set and RCC Regions

In relating the approximation regions of a single rough set separately with the two regions in an RCC-8 relation, it is easy to see that only five of the RCC-8 relations are possible: TPP(X,Y), NTPP(X,Y), EQ(X,Y), TPPI(X,Y), and NTPPI(X,Y). This follows since in a rough set the lower approximation region must be a subset of the upper approximation region. This condition does not hold true for the relations PO(X,Y), EC(X,Y), or DC(X,Y). Also, note that for the EQ(X,Y) relation, the upper and lower approximation regions are equal, resulting in a crisp set having no uncertainty. We are not interested, however, in simply combining two crisp regions in a relationship and expressing those as a single rough set. What we are interested in are the spatial relationships occurring between two vague regions, each represented as a rough set.

RCC Spatial Relationships Modeled as Relationships of Two Rough Sets

Let us now consider the RCC-8 relations in terms of two vague regions denoted by rough sets X and Y defined on some approximation space. The determination of whether or not a relationship holds for these two rough regions can also be expressed in rough set notation. In the discussion that follows, the RCC-8 relations will be expressed in terms of rough set approximation regions. In [20] the 46 relationships are listed along with “possible” representation of each of the RCC-8 relationships. The word “possible” is used here because there is more generality present than when using rough sets to express the relationships. This is because the approximation regions of rough sets, and therefore the vague boundaries, are each precisely defined based on some precisely defined equivalence relation. Vagueness in the egg yolk model includes uncertainty even in the specification of the boundaries of the egg and yolk. These boundaries “represent conservatively defined limits on the possible ‘complete crispings’ of a vague region” [20]. In addition to defining vague regions, rough sets can be used to describe RCC relationships. These relationships can be expressed in terms of which properties *certainly* hold and which *possibly* hold.

Recall the DC relationship represents DisConnected in RCC-8 theory. In rough set terms, the DC(X,Y) relationship holds when the rough sets X and Y are disconnected. This is *certainly* true when

$\overline{R}X \cap \overline{R}Y = \emptyset$, Figure 6, case 1. However, it is possible that the relation holds when $\underline{R}X \cap \underline{R}Y = \emptyset$ and $\overline{R}X \cap \overline{R}Y \neq \emptyset$. This occurs in several cases, some of which may also be found in Figure 6.

For the EQ(X,Y) relationship to hold certainly, $\underline{R}X = \underline{R}Y$ and $\overline{R}X = \overline{R}Y$. The EQ relationship possibly holds when $\underline{R}X = \underline{R}Y$ and either $\overline{R}X \subset \overline{R}Y$ or $\overline{R}Y \subset \overline{R}X$, which includes the two additional relationships that each have the two lower approximations equal and one upper approximation contained within the other. There are about 18 possibilities because the restriction on the inner boundaries being equal is relaxed. For rough sets, where this inner boundary is defined by the lower approximation region, equality must hold.

The relations TPP(X,Y) and NTPP(X,Y) are indistinguishable in rough set theory since the regions can be related by the subset relationship, but not quantified for intersection at one and only one point. Nor do we separate those points on the outer edge enclosing a region from those points in the interior of a region. They all equally belong to the region. TPP(X,Y) and NTPP(X,Y) together denote inclusion of rough set X in rough set Y. In rough set terms, $X \subseteq Y$ when $\underline{R}X \subseteq \underline{R}Y$ and $\overline{R}X \subseteq \overline{R}Y$. The relations TPPI(X,Y) and NTPPI(X,Y) are analogous to TPP(X,Y) and NTPP(X,Y). One may simply exchange the X and Y in the relation and discussion for TPP or NPP to obtain the same results. Grouped together these four relations certainly hold for 9 samples and possibly hold for Sample 46. All but five of the pairs may be categorized as having this relationship, which makes sense if we consider that we are looking at how two vague regions might relate to each other, this relationship basically interpreted as one region being “covered by” another.

Partial overlap PO(X,Y) implies that X and Y are not equal, but that they have some part in common. Rough set expression of this relationship involves both intersection and equality. This relationship will certainly hold whenever $\underline{R}X \cap \underline{R}Y \neq \emptyset$. It will possibly hold when $\overline{R}X \cap \overline{R}Y \neq \emptyset$. These results are identical to the 41 samples in [20] that meet the requirements for partially overlapping.

Because rough set theory does not allow us to specifically denote that two rough sets intersect at exactly one point, the EC(X,Y) relationship does not apply. It expresses the same relationship as those belonging to the possible region for the PO relationship discussed previously.

It is evident from the discussion above that rough sets can be used to model relationships for vague regions defined by the RCC-5 method. By allowing the expression of belonging to the relationship to be either certain or possible, however, we gain greater insight into the relationship, therefore greater knowledge. Rough sets also provide the indiscernibility relation, which aids in quantifying the uncertainty through the approximation regions. We now consider another approach for vague regions.

Egg-Yolk Approach

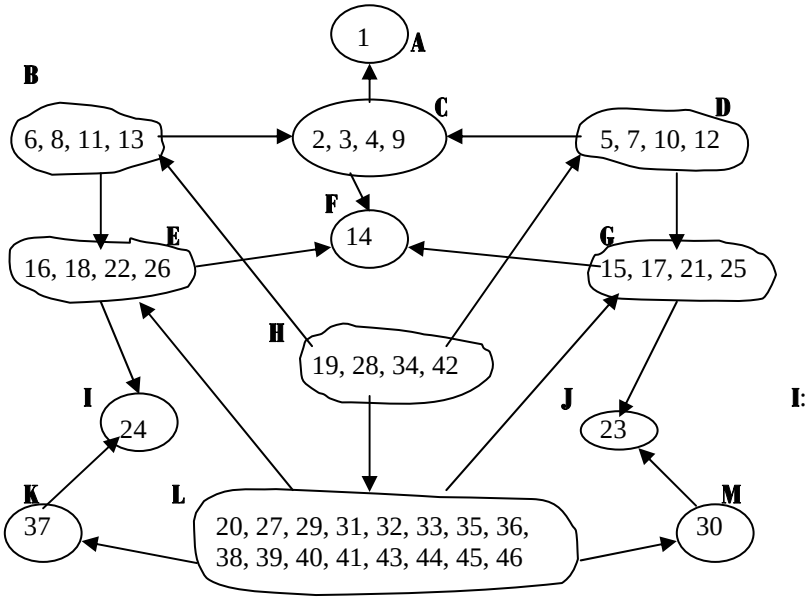
If we are only concerned about the vagueness of boundaries, we may be inclined to use the egg-yolk approach. In this approach concentric subregions make up a vague region, with inner subregions having the property that they are 'crisper,' or less imprecise, than outer subregions. These regions indicate a type of membership in the vague region. The simplest case, is that of two subregions. In this most common case, the center region is known as the yolk, the outer region surrounding the yolk is known as the white, and the entire region, as the egg. Although the boundaries of these subregions are also vague, the yolk and egg represent limits on the boundary of the vague region, which actually contains an infinite number of regions falling between these yolk and egg borders.

Consider how the yolk and egg compare to the boundary regions of rough sets. The rough set theory has only these two approximation regions, unlike the possible numerous subregions that may make up a vague region in the egg-yolk method. However, because of the indiscernibility relation in rough sets, one can vary the partitioning in order to increase or decrease the level of uncertainty present, which provides us with a formal mechanism to tune changes to the approximation regions. A finer partitioning results in a crisper region, one having less imprecision.

Let us now consider specifically the results of Cohn and Gotts [20]. In their paper they delineate 46 possible egg-yolk pairs (see Figure 3 for a representative sample), showing all of the possible relationships between two vague regions. They then relate the egg-yolk configurations to dyadic relations

of the type C(X,Y), meaning “C connects X with Y”, of their RCC-5 theory of spatial regions. Pairs of relations are called “immediate conceptual neighbors” if “each can be transformed into the other by a process of gradual, continuous change that does not involve passage through any third relation.” Recall that the RCC-5 relations include DR (Distinct Regions with no overlap), EQ (Equal: the regions are the same), PO (the regions Partially Overlap), PP (Proper Part: the first region is entirely contained within the second), and PPI (Proper Part Inverse: the first region entirely contains the second). For egg-yolk pairs a yolk is a PP of its own egg.

The 46 configurations of egg-yolk pairs were then clustered into 13 groups based on RCC-5 relations between complete crispings, or relations that are “mutually crispable” (see Figure 7). A complete crisping results when imprecision is reduced in some fashion until it reaches a point where there is no longer any imprecision. The clusters each relate to one or more additional clusters via a crisping relationship or a subset relationship between a set of complete crispings. Each configuration of a cluster can be crisped to the cluster pointed to by the arrow via one of these relationships. Notice the symmetrical nature of the figure. This results when X and Y are “reversed” in a relationship, such as the relationships grouped in clusters B and D, or in I and J.



RCC-5 Relations for Cluster

A:	DR
B:	DR, PO, PP
C:	DR, PO
D:	DR, PO, PPI
E:	PO, PP
F:	PO
G:	PO, PPI
H:	DR, PO, PP, PPI, EQ
I:	PP
J:	PPI
K:	EQ, PP
L:	EQ, PO, PP, PPI
M:	EQ, PPI

Figure 7.
Clustering of the 46 relations using RCC-5 relations between complete crispings of the configurations.

Rough Set Approach to Egg-Yolk Clustering

We will now re-examine the clustering of egg-yolk pairs, this time noting the relationships for each cluster based on formal definitions involving rough sets. Recall that "crisping" from the egg-yolk theory can be related to forcing a finer partitioning on the domain for rough sets.

We first review a few of definitions from rough set theory to be used in categorizing the clusters:

Two rough sets X and Y are equal, X = Y, if $\underline{R}X = \underline{R}Y$ and $\overline{R}\,X = \overline{R}\,Y$.

The intersection of two rough sets is defined by the approximation regions as follows:

$$\underline{R}(X \cap Y) = \underline{R}X \cap \underline{R}Y, \text{ and } \overline{R}\,(X \cap Y) = \overline{R}\,X \cap \overline{R}\,Y.$$

The subset relationship, $X \subset Y$ implies that $\underline{R}X \subset \underline{R}Y$ and $\overline{R}\,X \subset \overline{R}\,Y$.

Rough Set Characterization of Clusters

Now look again at Figure 7, but this time approach the clusters in terms of rough sets instead of RCC-5 relations. Let X denote the first egg (shown by dashed line) and Y denote the second egg (denoted by dotted line) in each egg pair, as shown in Figure 6. We then have the following characterizations for the clusters:

A: $X \cap Y = \emptyset$.

B: $\underline{R}X \cap \underline{R}Y = \emptyset \ \wedge \ \overline{R}\,X \cap \overline{R}\,Y \neq \emptyset \wedge \ \underline{R}X \subset \overline{R}\,Y$.

C: $\underline{R}X \cap \underline{R}Y = \emptyset \wedge \overline{R}\,X \cap \overline{R}\,Y \neq \emptyset$.

D: $\underline{R}X \cap \underline{R}Y = \emptyset \wedge \overline{R}\,X \cap \overline{R}\,Y \neq \emptyset \wedge \underline{R}Y \subset \overline{R}\,X$.

E: $\underline{R}X \cap \underline{R}Y \neq \emptyset \wedge \ \overline{R}\,X \cap \overline{R}\,Y \neq \emptyset \wedge \ \underline{R}X \cap \overline{R}\,Y \neq \emptyset \wedge \ \overline{R}\,X \cap \underline{R}Y \neq \emptyset \wedge \ \underline{R}X \subset \overline{R}\,Y$.

F: $\underline{R}X \cap \underline{R}Y \neq \emptyset \ \wedge \ \overline{R}\,X \cap \overline{R}\,Y \neq \emptyset \ \wedge \ \underline{R}X \cap \overline{R}\,Y \neq \emptyset \wedge \ \overline{R}\,X \cap \underline{R}Y \neq \emptyset \ \wedge \ \underline{R}X \subset \overline{R}\,Y$
$\wedge \ \underline{R}X \not\subseteq \overline{R}\,Y \ \wedge \ \underline{R}Y \not\subseteq \overline{R}\,X$

G: $\underline{R}X \cap \underline{R}Y \neq \emptyset \ \wedge \ \overline{R}\,X \cap \overline{R}\,Y \neq \emptyset \ \wedge \ \underline{R}X \cap \overline{R}\,Y \neq \emptyset \ \wedge \ \overline{R}\,X \cap \underline{R}Y \neq \emptyset \wedge \ \underline{R}Y \subset \overline{R}\,X$.

H: $\underline{R}X \subset \ \overline{R}\,Y \wedge \ \underline{R}Y \subset \overline{R}\,X \ \wedge \underline{R}X \cap \underline{R}Y \neq \emptyset$

I: $\overline{R}\,X \subset \underline{R}Y$.

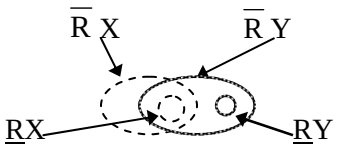
J: $\overline{R}\,Y \subset \underline{R}X$.

K: $\overline{R}\,X = \underline{R}Y$.

L: $\underline{R}X \cap \underline{R}Y \neq \emptyset \ \wedge \ \underline{R}X \subset \overline{R}\,Y \wedge \ \underline{R}Y \subset \overline{R}\,X \ \wedge \ \overline{R}\,X \cap \overline{R}\,Y \neq \emptyset$.

M: $\overline{R}\,Y = \underline{R}X$.

We can obtain the above results by examination of each case. Also we can express the relation between RCC relations and rough sets. Consider the result for cluster B. Examining again instance 6 of Figure 6 we have



Clearly $\underline{R}X$ and $\underline{R}Y$ do not overlap, so we have the first condition $\underline{R}X \cap \underline{R}Y = \varnothing$. Also note that $\overline{R}X$ and $\overline{R}Y$ do overlap, and that $\underline{R}X$ is completely contained in $\overline{R}Y$ and so we then have the last two conditions $\overline{R}X \cap \overline{R}Y \neq \varnothing$ and $\underline{R}X \subset \overline{R}Y$. These three conditions suffice to describe the configuration of instance 6 (Figure 6). They are the most general conditions that all the cases in cluster B satisfy. We shall discuss further properties of this cluster in more detail shortly. We can obtain all the results by examination of each case in clusters A through M as above. Also we can derive the results by considering the relationships between RCC relationships for each cluster and rough set relationships.

Consider, for example, cluster M. This cluster contains only the egg/yolk pair sample 30, which can be found in Figure 6. This cluster is based on the RCC relations EQ and PPI, which means that the regions might be equal or that region X entirely contains region Y. In rough set terminology, this relationship is defined through the use of the approximation regions. The relationship holds true whenever the upper approximation region of Y is equal to the lower approximation of X. We know certainly then, that Y is contained in X. It is also possible, however, that X and Y are the same (equal).

Let us consider our example of cluster B, which includes the egg-yolk pair relationships shown in Figure 8 below, also in this manner. The RCC-5 relations corresponding to this cluster include DR, PO, and PP. We can see that for these relationships it is always the case that the two lower approximation regions do not intersect ($\underline{R}X \cap \underline{R}Y = \varnothing$), DR, but that the two upper approximation regions do intersect ($\overline{R}X \cap \overline{R}Y \neq \varnothing$), PO. These properties also hold true for the relationships in cluster D, which is

symmetric with cluster B in terms of X and Y. For cluster B, it is also true that the lower approximation region of X is contained in the upper approximation region of Y ($\underline{R}X \subset \overline{R}\,Y$), PP. In D, the relationship is reversed, with $\underline{R}Y \subset \overline{R}\,X$, PPI.

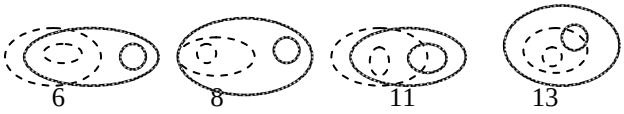


Figure 8. Relationships between regions X (dashed line) and Y (dotted line) found in cluster B of Figure 7.

SubClustering Characterizations

Let us now examine cluster B in more detail. We shall see that the structure of the cluster and others can be refined by a rough set analysis. In both the clustering for egg-yolk and 9-intersection methods [20], we can have the group arranged hierarchically as depicted in Figure 9a. This cluster denotes the relationship Close, "X is close to Y". Notice that in a sense sample 6 is the most general and "least close," whereas sample 13 is the "most close". To transform the relationship from sample 6 to sample 8, one of the upper approximation regions is increased or decreased so that the upper approximation of X is entirely contained in the upper approximation of Y. To transform the relationship from sample 6 to sample 11, we can either change the upper approximation of X or the lower approximation of Y so that these two intersect. However in order to obtain the relationship in sample 13 from that of sample 6, we must proceed through both of these operations, and so 6 and 13 are not contained in the same subcluster.

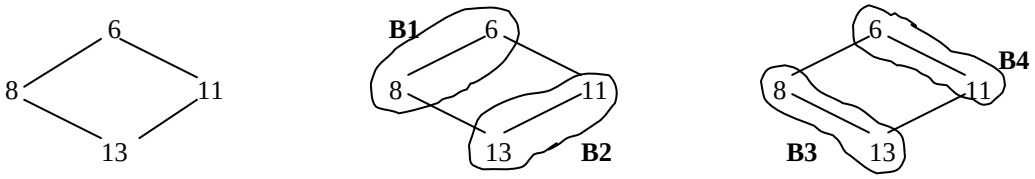


Figure 9. (a.) Hierarchical property of cluster B. (b.) Clustering B1, B2. (c.) Clustering B3, B4.

Subclusters for B and D

As discussed previously, the rough set terminology for representation of the particular cluster B can be expressed by the following:

B: $\underline{R}X \cap \underline{R}Y = \varnothing \wedge \overline{R}X \cap \overline{R}Y \neq \varnothing \wedge \underline{R}X \subset \overline{R}Y.$

With rough sets, however, this cluster can be further decomposed into two smaller clusters in two different ways. The first subclustering groups together 6 and 8 into subcluster B1 by including the additional property $\underline{R}Y \cap \overline{R}X = \varnothing$ valid for both:

B1: $\underline{R}X \cap \underline{R}Y = \varnothing \wedge \overline{R}X \cap \overline{R}Y \neq \varnothing \wedge \underline{R}X \subset \overline{R}Y \wedge \underline{R}Y \cap \overline{R}X = \varnothing,$

and 11 and 13 into another cluster B2 having the complementary property $\underline{R}Y \cap \overline{R}X \neq \varnothing$ (Figure 9b):

B2: $\underline{R}X \cap \underline{R}Y = \varnothing \wedge \overline{R}X \cap \overline{R}Y \neq \varnothing \wedge \underline{R}X \subset \overline{R}Y \wedge \underline{R}Y \cap \overline{R}X \neq \varnothing.$

We might say that the two regions, because of the non-null intersection, are in a sense "closer" in the subcluster B2 than in B1.

We can form yet another subclustering of cluster B by grouping together 8 and 13 in subcluster B3 and 6 and 11 in subcluster B4 (Figure 9c). In B3, the property $\overline{R}X \subset \overline{R}Y$ is added:

B3: $\underline{R}X \cap \underline{R}Y = \varnothing \wedge \overline{R}X \cap \overline{R}Y \neq \varnothing \wedge \underline{R}X \subset \overline{R}Y \wedge \overline{R}X \subset \overline{R}Y.$

Here Y possibly "surrounds" X. In B4, however, we add the property $\overline{R}X \not\subseteq \overline{R}Y$.

B4: $\underline{R}X \cap \underline{R}Y = \varnothing \wedge \overline{R}X \cap \overline{R}Y \neq \varnothing \wedge \underline{R}X \subset \overline{R}Y \wedge \overline{R}X \not\subseteq \overline{R}Y.$

Both subclusters retain the original properties of B as well.

Because the cluster D

D: $\underline{R}X \cap \underline{R}Y = \varnothing \wedge \overline{R}X \cap \overline{R}Y \neq \varnothing \wedge \underline{R}Y \subset \overline{R}X$

contains 5, 7, 10, and 12 which are the mirror images of 6, 8, 11, and 13, with X and Y reversed, we can likewise form 2 possible subclustering of this group denoting "Y is close to X". The first subclustering groups

5 and 7 together (D1) since $\underline{R}X \cap \overline{R}\, Y = \varnothing$, and 10 and 12 together (D2), with the complementary property yielding:

D1: $\underline{R}X \cap \underline{R}Y = \varnothing \wedge \overline{R}\, X \cap \overline{R}\, Y \neq \varnothing \wedge \underline{R}Y \subset \overline{R}\, X \wedge \underline{R}X \cap \overline{R}\, Y = \varnothing$ and
D2: $\underline{R}X \cap \underline{R}Y = \varnothing \wedge \overline{R}\, X \cap \overline{R}\, Y \neq \varnothing \wedge \underline{R}Y \subset \overline{R}\, X \wedge \underline{R}X \cap \overline{R}\, Y \neq \varnothing$

The second subclustering groups 7 and 12 together (D3) by containment, ($\overline{R}\, Y \subset \overline{R}\, X$) and then 5 and 10 by using the complement:

D3: $\underline{R}X \cap \underline{R}Y = \varnothing \wedge \overline{R}\, X \cap \overline{R}\, Y \neq \varnothing \wedge \underline{R}Y \subset \overline{R}\, X \wedge \overline{R}\, Y \subset \overline{R}\, X$ and
D4: $\underline{R}X \cap \underline{R}Y = \varnothing \wedge \overline{R}\, X \cap \overline{R}\, Y \neq \varnothing \wedge \underline{R}Y \subset \overline{R}\, X \wedge \overline{R}\, Y \not\subseteq \overline{R}\, X$.

We see, therefore, that rough sets can be used to naturally express a grouping of relationships that are not expressed in the original clustering of B and D found in [11].

Subclusters for H and C; E and G

Now we can see that subclusters can be formulated for several other original clusters as shown in the previous section for B and D. Note that the general approach was to first identify a subclustering property P. Then P and its complement P* were used to define subclusters, and this will be the general approach for the following discussion. Finally we will provide an overall summarization of the subclustering results in Table 1.

First let us consider the cluster H whose instance are shown in Figure 10. These relations are grouped together for egg-yolk relations, but are not included at all in the clustering scheme based on the 9-intersection model since the assumption was made that the indeterminate region was very small in comparison with the entire region. We can arrange the four relations hierarchically as shown in Figure 11 and subcluster through additional rough set properties. As with the hierarchical subclustering done for clusters B and D, note how sample 42 of cluster H appears to be the "least imprecise" and sample 19, the "most imprecise," with samples 28 and 34 having some level of precision between 19 and 42, all based on the relationships between upper approximation regions only for this cluster.

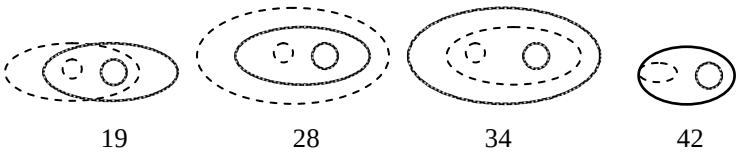


Figure 10. Relationships between regions X (dashed line) and Y (dotted line) found in cluster H of Figure 7.

We may again form subclusters of this group in two ways. The first subclustering groups together relations 34 and 42 in one cluster (H1) and 19 and 28 in one cluster (H2). The rough set property P defining these subcluster is $\overline{R}\,X \subseteq \overline{R}\,Y$ producing:

H1: $\underline{R}X \subset (\overline{R}\,X \cap \overline{R}\,Y) \wedge \underline{R}Y \subset (\overline{R}\,X \cap \overline{R}\,Y) \wedge \overline{R}\,X \subseteq \overline{R}\,Y.$
H2: $\underline{R}X \subset (\overline{R}\,X \cap \overline{R}\,Y) \wedge \underline{R}Y \subset (\overline{R}\,X \cap \overline{R}\,Y) \wedge \overline{R}\,X \not\subseteq \overline{R}\,Y,$ and

Notice that in contrast to the conditions for subclusters in B and D in which we used the stronger condition of a proper subset, here we use $\subseteq$. The reason is that in fact, for sample 42 we have $\overline{R}\,X = \overline{R}\,Y$ as seen in Figure 10.

A different clustering may be obtained analogously by interchanging X and Y in P giving:

H3: $\underline{R}X \subset (\overline{R}\,X \cap \overline{R}\,Y) \wedge \underline{R}Y \subset (\overline{R}\,X \cap \overline{R}\,Y) \wedge \overline{R}\,Y \subseteq \overline{R}\,X$ and
H4: $\underline{R}X \subset (\overline{R}\,X \cap \overline{R}\,Y) \wedge \underline{R}Y \subset (\overline{R}\,X \cap \overline{R}\,Y) \wedge \overline{R}\,Y \not\subseteq \overline{R}\,X$

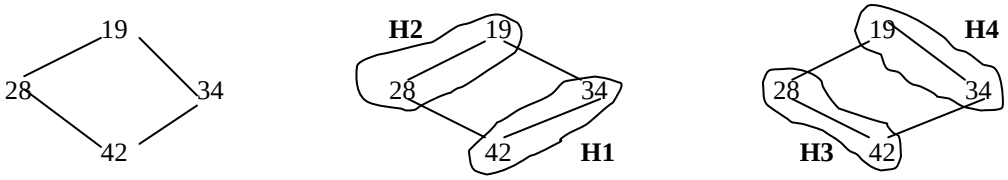


Figure 11. (a.) Hierarchical property of cluster H. (b.) H1 and H2. (c.) H3 and H4

We may use similar techniques on cluster C (contains relations 2, 3, 4, and 9) to transform it into two different subclusterings of two relations each. The first partitioning clusters 2 , 3, 4 and 9 based on the property $\underline{R}X \cap \overline{R}\,Y$. So the subclusters C1 and C2 have properties of C along with those below:

C1: $\underline{R}X \cap \underline{R}Y = \varnothing \wedge \overline{R}\,X \cap \overline{R}\,Y \neq \varnothing \wedge \underline{R}X \cap \overline{R}\,Y = \varnothing,$ and
C2: $\underline{R}X \cap \underline{R}Y = \varnothing \wedge \overline{R}\,X \cap \overline{R}\,Y \neq \varnothing \wedge \underline{R}X \cap \overline{R}\,Y \neq \varnothing.$

As in the second subclustering for H described above, interchanging X and Y in the first subclustering property yields a new subclustering. So here there is another subclustering of C, based on the property $\overline{R}\,X \cap \underline{R}Y = \varnothing$, forming subcluster C3 with 2 and 4, and a second subcluster C4, containing 3 and 9.

Finally let us consider Cluster E

E: $\underline{R}X \cap \underline{R}Y \neq \varnothing \wedge \overline{R}\,X \cap \overline{R}\,Y \neq \varnothing \wedge \underline{R}X \cap \overline{R}\,Y \neq \varnothing \wedge \overline{R}\,X \cap \underline{R}Y \neq \varnothing \wedge \underline{R}X \subset \overline{R}\,Y$ containing 16, 18, 22, and 26 and depicted in Figures 12 and 13 below, (and likewise its mirror image cluster G) which can also be divided into subclusters in two ways. The first grouping places 16 and 18 into one cluster (E1), based on the relationship between the upper approximations of X and Y, $\overline{R}\,X \not\subseteq \overline{R}\,Y$, and 22 and 26 in the other cluster (E2) with the property $\overline{R}\,X \subset \overline{R}\,Y$.

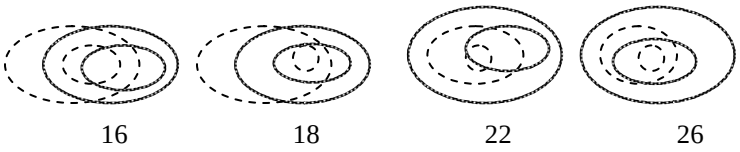


Figure 12. Relationships between regions X (dashed line) and Y (dotted line) found in cluster E of Figure 7.

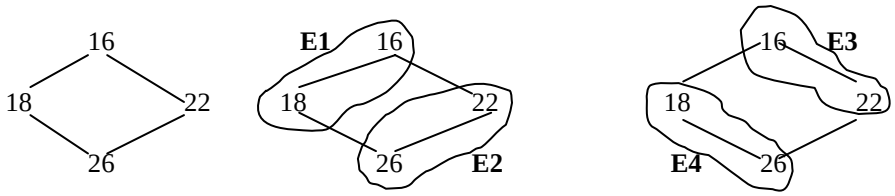


Figure 13. (a.) Hierarchical property of cluster E. (b.) Clustering E1, E2. (c.) Clustering E3, E4.

A different clustering of E is obtained based on the subset relationship of the lower approximations of X, grouping 16 and 22 in one cluster (E3) and 18 and 26 in the other (E4). For E3, $\underline{R}X \not\subseteq \underline{R}Y$, and for E4, $\underline{R}X \subset \underline{R}Y$.

Cluster E denotes the spatial relation "covered by". The hierarchy in Figure 13 arises because in 26 this "covered by" relation is "more certain" and in 16 it is "less certain" than the others. Relations 18 and 22 fall somewhere between 16 and 26 in their level of certainty. Cluster G can be analyzed and subclustered analogously to cluster E by simply exchanging X and Y, 16 with 15, 22 with 21, 26 with 25, and 18 with 17 for similar results.

Let us provide a summary of these results. There are four distinct properties producing subclustering:

P_1: $\underline{R}Y \cap \overline{R}\,X = \varnothing$

P_2: $\overline{R}\,X \subseteq \overline{R}\,Y$

P_3: $\underline{R}X \subset \underline{R}Y$

P_4: $\overline{R}\,X \subset \overline{R}\,Y$

For each property $P_i(X, Y)$, the property $P_i(Y, X)$ represents the property with X and Y interchanged, e.g. $P_2(Y, X)$: $\overline{R}\,Y \subseteq \overline{R}\,X$. Now we can develop Table 1 which provides the results for the subclustering relative to their defining properties.

Table 1. Summary of Subclusters.

PROPERTY	$P_i(X, Y)$	$P_i(Y, X)$
P_1	B1-2, C3-4	D1-2, C1-2
P_2	B3-4, E1-2	D3-4, G1-2
P_3	H1-2	H3-4
P_4	E3-4	G3-4

The previous discussion illustrates the expressive power of rough sets and the generality of the approximation regions of rough set theory in formalizing relationships for vague regions. We have shown that rough sets can be used to express all the spatial relationships defined for 9-intersection, RCC, and egg-yolk methods. In addition we have given several examples of subclustering that can be expressed in

terms of rough set properties. Recall that rough sets offer the additional capability of partitioning through the use of the equivalence relation. Rough set techniques for information retrieval and data mining in spatial databases and geographic information systems (GIS) may be applied to models incorporating vague regions expressed by the egg-yolk and RCC models, as well as those based on the 9-intersection method. Rough set theory provides a comprehensive mathematical foundation for vague regions that is compatible with other spatial data theories and methods, yet offers the added ability to refine the indiscernibility.

Conclusion

Spatial and geographical information systems will continue to play an ever-increasing role in applications based on spatial data. Uncertainty management will be necessary for any of these applications, and rough sets, 9-intersection, RCC, and egg-yolk methods are appropriate for the representation of vague regions in spatial data. Rough sets, however, can also model indiscernibility and allow for the change of granularity of the partitioning through its indiscernibility relation. Changing the indiscernibility relation has an effect on the boundaries of the vague regions because the lower and upper approximation regions are defined in terms of this indiscernibility.

As discussed in [7], extending the 9-intersection model to relations between objects with broad boundaries maintains all the properties of the original 9-intersection model, giving a mutually exclusive set of relations, and providing an algebraic basis for spatial reasoning. It can easily be implemented in a GIS since each region of a broad boundary can be expressed as two sharp boundaries with ordinary polygons. An equivalent rough set representation has similar benefits.

Rough set techniques can also be used to define the spatial relationships themselves. In this manner, there is a distinction between those combinations that certainly meet the RCC-8 relationship requirements and those that possibly meet the requirements. The rough set approach, therefore, is very useful in defining vague spatial regions with indeterminate boundaries, and in defining the spatial relationships that hold between two vague regions.

We have also shown how the clustering of egg-yolk pairs by RCC-5 relations can be expressed in terms of operations using rough sets. We believe that rough set techniques can further enhance the egg-yolk approach and are investigating the interrelationships between rough set, egg-yolk, RCC, complex objects (region having holes and multiple components) [28] and other spatial models [29]. We are also investigating impact of vagueness and uncertainty expressed with these theories on the querying and mining of spatial data [30,31], and the feasibility of implementing such approaches [32].

ACKNOWLEDGEMENTS

We would like to thank the Naval Research Laboratory's Base Program, Program Element No.0602435N for sponsoring this research.

References

[1] P. Rigaux, M. Scholl and A. Voisard, *Spatial Databases: With Application to GIS*, Morgan Kaufmann Publishers, 2001.

[2] S. Shekhar and S. Chawla, *Spatial Databases: A Tour*, Prentice Hall, 2002.

[3] M. Goodchild and S. Gopal (eds), *Accuracy of Spatial Databases*, Taylor and Francis, 1989.

[4] P. Fisher, "Sorites Paradox and Vague Geographies", *Fuzzy Sets and Systems*, 113(1), 7-18, 2000.

[5] M Duckham,., K.Mason , J. Stell, and M.F.Worboys, "A formal ontological approach to imperfection in geographic information," *Computers, Environment and Urban Systems*, **25**: 89-103, 2001.

[6] T. Beaubouef and F. Petry , and B. Buckles, "Extension of the Relational Database and its Algebra with Rough Set Techniques," *Computational Intelligence*, Vol. 11, No. 2, May 1995, pp. 233-245.

[7] T. Beaubouef and F. Petry , "Rough Querying of Crisp Data in Relational Databases," *Third Int. Workshop on Rough Sets and Soft Computing (RSSC'94)*, San Jose, California, November 1994.

[8] T. Beaubouef, F. Petry, and J. Breckenridge, "Rough Set Based Uncertainty Management for Spatial Databases and Geographical Information Systems," in *Soft Computing in Industrial Applications* (ed. Y. Suzuki), Springer-Verlag, London, 2000.

[9] V. Robinson, "On Fuzzy Sets and the Management of Uncertainty in an Intelligent Geographic Information System", *Recent Issues on Fuzzy Databases* editors G. Bordogna, G. Pasi, , pp.109-128, Physica-Verlag, Heidelberg, GR, 2000.

[10] G. Bordogna, S. Chiesa and D. Geneletti, "Linguistic Modeling of Imperfect Spatial Information in a Fuzzy Database, in *Proc 9th IPMU 2002*, pp. 7-14, Ancy, Fr, 2002.

[11] H. Guesgen and J. Albrecht, “Imprecise Reasoning in Geographic Information Systems” *Fuzzy Sets and Systems*, 113(1), 121-132, 2000.

[12]M.F Worboys., “Imprecision in finite resolution spatial data,” *Geoinformatica* **2**(3): 257-280, 1998.

[13] T. Bittner and J.Stell,. “Vagueness and rough location.” *Geoinformatica*, vol. 6, pp. 99-121. 2002.

[14] Z.Pawlak, “Rough Sets,” *Int. Journal of Man-Machine Studies*, vol. 21, pp. 127-134, 1984.

[15] M. Egenhofer, and J. Herring, “Categorizing Binary Topological Relationships Between Regions, Lines, and Points in Geographic Databases,” Technical Report, University of Maine, Department of Surveying Engineering, 1991.

[16] E. Clementini, and P. di Felice, “An Algebraic Model for Spatial Objects with Indeterminate Boundaries,” in *Geographic Objects with Indeterminate Boundaries* (ed. P. Burrough and A. Frank), GISDATA II, European Science Foundation, chapter 11, pp. 155-169, 1996.

[17] D. Randell, and A.Cohn “Modeling topological and metrical properties of physical processes,” *Proc. First Int. Conf. On the Principles of Knowledge Representation and Reasoning*, Los Altos, 1989, pp. 55-66.

[18] Randell, D., Cui, Z., and Cohn, A., “A spatial logic based on regions and connection,” *Proc. 3rd Int. Conf. On Knowledge Representation and Reasoning*, San Mateo, 1992, pp. 165-176,

[19] F. Lehmann, and A. Cohn, “The EGG/YOLK reliability hierarchy: Semantic data integration using sorts with prototypes,” *Proc. 3rd Int. Conf. on Information and Knowledge Management*, Gaithersburg, MD, 1994, pp. 272-279.

[20] A. Cohn and Gotts, N. “The ‘Egg-Yolk’ Representation of Regions with Indeterminate Boundaries,” in *Geographic Objects with Indeterminate Boundaries* (ed. P. Burrough and A. Frank), GISDATA II, European Science Foundation, chapter 12, 1996.

[21] Z. Pawlak, *Rough Sets: Theoretical Aspects of Reasoning about Data*, Kluwer Academic Publishers, Norwell, MA., 1991.

[22] Komorowski, J., Pawlak, Z., Polkowski, L., et. al., “Rough Sets: A Tutorial,” in *Rough Fuzzy Hybridization: A New Trend in Decision-Making* (ed. S. K. Pal and A. Skowron), Springer-Verlag, Singapore, pp. 3-98, 1999.

[23] R.Slowinski, "A Generalization of the Indiscernibility Relation for Rough Sets Analysis of Quantitative Information," *First Int. Workshop on Rough Sets*, Poland, 1992.

[24] P Srinivasan, “The importance of rough approximations for information retrieval,” *International Journal of Man-Machine Studies*, 34, pp. 657-671, 1991.

[25] M.F. Worboys, and E.Clementini, “Integration of imperfect spatial information,” *Journal of Visual Languages and Computing*, **12**, 61-80, 2001.

[26] A. Roy and J. Stell, “Spatial Relations Between Indeterminate Regions,” *Int. Jour. Approximate Reasoning*, Vol. 27, No. 3, pp. 205-234, 2001.

[27] Chung, M, Cobb M, Foley H, Petry F, and Shaw K, "A Rule-based Approach for the Conflation of Attributed Vector Data”, *GeoInformatica*, 2, #1, pp. 1-29, 1998.

[28] E. Clementini and P. DiFelice, “A Spatial Model for Complex Objects with a Broad Boundary Supporting Queries on Uncertain Data”, *Data And Knowledge Engineering*, 37, pp. 285-305, 2001.

[29] M. Worboys, “Imprecision in Finite Resolution Spatial Data”, *Geoinformatica* **2**(3): 257-280, 1998

[30] E. Clementini, P. DiFelice and K. Koperski, “Mining Multiple-level Spatial Association Rules for Objects with a Broad Boundary, *Data And Knowledge Engineering*, 34, pp. 251-270, 2000.

[31] H. Miller and J. Han (eds.), *Geographic Data Mining and Knowledge Discovery*, Taylor and Francis, 2001

[32] T. Beaubouef and F. Petry , “A Rough Set Foundation for Spatial Data Mining Involving Vague Regions", *Proc FUZZ-IEEE 2002*, pp. 767-772, Honolulu, HW, 2002.